TR 659

Estimation of Tail Probability via the Maximum L_q-Likelihood method

Davide Ferrari

e-mail: dferrari@stat.umn.edu

and

Yuhong Yang

e-mail: yyang@stat.umn.edu

Abstract: Estimation of tail probability is of interest in various applications. Given a parametric model, a natural approach is maximum likelihood estimation. Although the resulting estimator is asymptotically efficient, the large sample property is often not trustworthy for estimating small tail probabilities. We introduce a new estimator for the parameters, called Maximum Lq-Likelihood Estimator (MLqE), based on Havrda and Charvát's entropy function (Havrda and Charvát, 1967), and apply it for estimating tail probabilities. The MLqE can be regarded as an extension of the traditional log-likelihood maximization procedure. Specifically, its behavior is characterized by the degree of distortion, q, applied to the assumed model; when q is close to 1 the new estimator approaches the usual MLE. We derive asympotic properties of the new estimator showing that when q is properly chosen according the sample size, the MLqE can successfully trade bias for variance when the amount of information available is not large relative to the size of the tail probability to be estimated. The case of the exponential distribution is considered, showing that when the distortion parameter q is properly chosen according to the sample size, the ratio of the asymptotic Mean Squared Error of MLqE over that of MLE is smaller than 1. Monte Carlo simulations are carried out to study the finite-sample performance of the estimator, confirming the theoretical findings.

1. Introduction

The importance of rare events is well known in different domains of natural and social sciences. For instance, extreme waves, rainfalls and floods are of basic importance in oceanography and hydrology; high wind speeds and extreme

imsart-generic ver. 2007/02/20 file: paper1_4.tex date: June 12, 2007

temperatures in meteorology. Large insurance losses, strong fluctuations in prices of bonds have an impact in economics. Occurrence of widespread and virulent epidemics are of great significance in epidemiology.

In modeling small tail probability, parametric classes of distributions can be employed. However, we must deal with the fact that in many applications the number of observations available is not large in relation to the occurrence of the event of interest. In such situations, traditional methods such as maximum likelihood might be severely affected by deviations from the chosen model [9], and inference can be arduous when the amount of information available is limited.

Robust versions of the classical likelihood method have been previously proposed to handle similar issues. Weighted likelihood (WL) maximization techniques have been successfully employed in different contexts when there is a need to reduce the role of some observations in order to trade the bias for precision (*e.g.*, see Hu and Zidek [7],[8]). The WL extends the local likelihood method of Tibshirani and Hastie [14] and it shares its underlying purpose with other methods such as weighted least squares and kernel smoothers which can reduce an estimator's variance while increasing its bias to reduce mean-squared error. In practice however, the advantages of weighted likelihood methods rely heavily on a proper selection of the weights, based on the data [16].

In this paper a modified version of the maximum likelihood method is proposed. We introduce the Maximum Lq-Likelihood esitimator (MLqE), which aims to improve the traditional maximum likelihood approach when estimating tail probability or a quantile. Our approach is ispired by a class of generalized measures of information, the α-order entropies [6], which have found extensive applications in physics, finance, biomedical sciences and other fields [4].

When estimating a fixed tail probability, standard large sample theory guarantees that the MLE is asymptotically efficient. Consequently, when the sample size is large, the MLE is at least as accurate as any other estimator. Nevertheless, for a moderate or small sample size, it turns out that our application of the generalized information measure can dramatically reduce the variance of the MLE at the expense of a slightly increased bias.

In our approach the role of the observations is alterated by slightly changing the the model of reference by means of a distortion parameter q. From this standpoint, L_q-likelihood estimation can be regarded as the minimization of the discrepancy of a general model with respect to one that emphasizes (or diminishes) the role of extreme observations. Specifically, the specific type of bias introduced allows to gain in terms of precision when both the sample size and the tail probability to be estimated are small. Conversely, when the sample size is large, reducing the amount of distortion allows for the recovery of a number of desirable properties such as consistency, efficiency and asymptotic normality.

The paper is organized as follows. In section 2, we examine some information-theoretical quantities and introduce the MLqE; in section 3 we discuss its basic asymptotic properties. In section 4, we consider the plug-in approach for estimating the tail probability. The asymptotic properties of the plug-in estimator are derived and its efficiency is compared to the traditional MLE. In section 5 we present a Monte Carlo simulation study on the case of an exponential distribution; we examine the behavior of MLqE in finite sample situation and compare its performance to that of MLE.

2. Generalized entropy and the Maximum L_q-Likelihood Estimator

The Kullback-Leibler (KL) divergence ([11],[10]), or relative entropy, is one of the most popular quantities employed to measure the distance of a "target" distribution with respect to a "true" distribution. Consider a measure space Ω, μ and let $\mathcal{M}$ be the set of all probability distribution functions (pdfs) f normalized w.r.t., μ, $\int_\Omega f(x)d\mu(x) = 1$. The expectations with respect to f are denoted E_f. The KL divergence between two density functions g and f is

$$\mathcal{D}(f||g) = E_f \log\left(\frac{f(X)}{g(X)}\right) = \int_\Omega \log\left(\frac{f(x)}{g(x)}\right) f(x)d\mu(x). \tag{2.1}$$

Note that finding the density g that minimizes $\mathcal{D}(f||g)$ is equivalent to minimizing Shannon's entropy [13]:

$$\mathcal{H}(f||g) = -E_f \log g(X). \tag{2.2}$$

Since KL divergence was introduced, other and more general measures of information have been developed. In the mid-60s and 70s, Rényi [12], Aczél and Daróczy [1] introduced a class of information measures by keeping the additivity of independent mean information, while employing a more general definition of mean (usually such information measures are referred to as Rényi entropies). In contrast, Havrda and Charvát [6] introduced the α-order entropies (or q-entropies in physics), where the usual definition of mean is maintained, while the additivity assumption is removed.

More recently, q-entropies have been of increasing interest in different domains. Tsallis [15] has successfully exploited distorted information measures in physics in relation to non-equilibrum phenomena. Since then, a considerable amount of applications have appeared in various disciplines such as finance, biomedical sciences, environmental sciences and linguistic (*e.g.*, see Gell-Mann [4]).

Definition 2.1. *Let f and g be two densitity functions; the q-entropy is defined as*

$$\mathcal{H}_q(f,g) = -E_f\left[L_q\left(g(X)\right)\right], \quad q > 0, \tag{2.3}$$

where

$$L_q(u) = \begin{cases} \frac{u^{1-q}-1}{1-q} & \text{if } q \neq 1, \\ \log u & \text{if } q = 1. \end{cases} \tag{2.4}$$

The function L_q represents a Box-Cox transformation in statistics and in other contexts it is often called deformed logarithm. The above characterization emphasizes the resemblance to the classical Shannon's entropy; if $q \to 1$, then $L_q(u) \to \log(u)$ and the usual definition of Shannon's entropy is recovered.

Let $\mathcal{M}(\theta)$ be a family of parametrized density functions and suppose that the "true" density, denoted by $f(x;\theta_0)$, is a member of $\mathcal{M}(\theta)$. Assume further that $\mathcal{F}(\theta)$ is closed with respect the transformation

$$f(x;\theta)^{(r)} = \frac{f(x;\theta)^r}{\int_\Omega f(x;\theta)^r d\mu(x)}, \; r > 0. \tag{2.5}$$

The transformation $f(x;\theta)^{(r)}$ is often referred as to *zooming* or *escort* distribution and the parameter r provides a tool to accentuate different regions of the untransformed true density $f(x;\theta)$. In particular, when $r < 1$ regions with density values close to zero are accentuated, while for $r > 1$ regions with density values further from zero are emphasized.

Next, consider the KL divergence between $f(x;\theta)$ and $f(x;\theta_0)^{(r)}$:

$$\mathcal{D}_r(\theta_0, \theta) = \int_\Omega \log\left(\frac{f(x;\theta_0)^{(r)}}{f(x;\theta)}\right) f(x;\theta_0)^{(r)} d\mu(x). \tag{2.6}$$

If we let θ^* be the value such that $f(x;\theta^*) = f(x;\theta_0)^{(r)}$ and assume that differentiation can be passed under the integral sign, we can write

$$\begin{aligned} \frac{\partial}{\partial\theta}\mathcal{D}_r(\theta_0||\theta)\bigg|_{\theta=\theta^*} &= -\int_\Omega \frac{\partial}{\partial\theta}\log f(x;\theta)\bigg|_{\theta=\theta^*} f(x;\theta_0)^{(r)} d\mu(x) && (2.7) \\ &= -\int_\Omega \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\bigg|_{\theta=\theta^*} f(x;\theta_0)^{(r)} d\mu(x) = 0 && (2.8) \end{aligned}$$

Moreover, it can be verified that $\frac{\partial^2}{\partial\theta^2}\mathcal{D}_r(\theta_0||\theta)\Big|_{\theta=\theta^*} > 0$; thus, $\theta = \theta^*$ is a minimum.

Furthermore, let θ^{**} be the value such that $f(x;\theta^{**}) = f(x;\theta_0)^{(1/q)}$, where q is a positive constant. Assuming the validity of differentiation under the integral sign, we have

$$\frac{\partial}{\partial\theta}\mathcal{H}_q(\theta_0,\theta)\Big|_{\theta=\theta^{**}} = -\int_\Omega \frac{\partial}{\partial\theta}L_q\left(f(x;\theta)\right)\Big|_{\theta=\theta^{**}} f(x;\theta_0)d\mu(x) \tag{2.9}$$

$$= -\int_\Omega \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)^q}\Big|_{\theta=\theta^{**}} f(x;\theta_0)d\mu(x) = 0 \tag{2.10}$$

It can be verified that $\frac{\partial^2}{\partial\theta^2}\mathcal{H}_q(\theta_0,\theta)\Big|_{\theta=\theta^{**}} > 0$ and $\mathcal{H}_q(\theta_0,\theta)$ has a minimum at θ^{**}.

The derivations above suggest that the task of minimizing the q-entropy can be regarded as equivalent to the minimization of the KL relative divergence between the true distribution and the escort distribution, when $q = r^{-1}$. Clearly, by considering the divergence with respect to a distorted version of the true density we introduce a certain amount of bias. Nevertheless, the bias can be promptly controlled by an adequate choice of the distortion parameter q, and later we shall discuss the benefits gained from paying such a price.

The next definition introduces the estimator based on the empirical version of the q-entropy.

Definition 2.2. *Let $X_1, ..., X_n$ be an i.i.d. sample from $f(x;\theta_0)$, $\theta_0 \in \Theta$. The Maximum L_q-Likelihood Estimator (MLqE) of θ_0 is defined as*

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta}\sum_{i=1}^n L_q\left[f(X_i;\theta)\right], \quad q > 0, \tag{2.11}$$

where L_q is the q-logarithmic function defined in (2.4) with $q > 0$. The L_q-likelihood equation is

$$\frac{\partial}{\partial\theta}\sum_{i=1}^n L_q\left[f(X_i;\theta)\right] = 0. \tag{2.12}$$

Note that when the distortion parameter q tends to 1, $L_q(\cdot) \to \log(\cdot)$ and the usual MLE is recovered. In this sense, the MLqE extends the classic method, resulting in a general inferential procedure that inherits most of the desirable features of traditional maximum likelihood, and at the same time gains some new properties that can be exploited *ad hoc* in particular estimation settings.

Example 2.1. The simple case of an exponential distribution will be used as a recurrent example in the course of the paper. Consider an i.i.d. sample of size n from

$$f(x;\lambda_0) = e^{-x\lambda_0+\log\lambda_0}, \quad x>0, \quad \lambda_0>0. \tag{2.13}$$

In this case the L_q-likelihood equation by Chris Franklin and Alan Agresti ... Instructors Solutions Manual by Alan Agresti and Barbara Finlay ... is

$$0 = \frac{\partial}{\partial\lambda}\sum_{i=1}^{n} L_q\left[e^{-X_i\lambda+\log\lambda}\right] = \sum_{i=1}^{n} e^{-[X_i\lambda-\log\lambda](1-q)}\left(-X_i+\frac{1}{\lambda}\right).$$

Note that setting $q=1$ gives the usual Maximum Likelihood equation and the solution of the equation above is $\hat{\lambda} = \left(\sum_i X_i/n\right)^{-1} = \overline{X}^{-1}$, the arithmetic average of the observations. However, when $q \neq 1$, equation (2.1) can be rewritten as

$$\lambda = \frac{\sum_{i=1}^{n} e^{-[X_i\lambda-\log\lambda](1-q)}}{\sum_{i=1}^{n} X_i e^{-[X_i\lambda-\log\lambda](1-q)}}. \tag{2.14}$$

In particular, by setting $c_i := e^{-[X_i\lambda-\log\lambda](1-q)}$ we have:

$$\lambda = \left(\frac{\sum_{i=1}^{n} X_i c_i\left(X_i,\lambda,q\right)}{\sum_{i=1}^{n} c_i\left(X_i,\lambda,q\right)}\right)^{-1}. \tag{2.15}$$

Eq. (2.15) offers a natural interpretation of the MLqE as function of a *weighted* average of the observations. In this case, we remark that when $q<1$ the role played by observations corresponding to higher density values are accentuated; on the other hand if $q>1$ observations corresponding density values close to zero are accentuated.

3. Exponential Families and Asymptotics of the MLqE

In this section, we discuss the basic asymptotic properties of the new estimator when the degree of distortion depends on the amount of information available in the sample. Such properties will be used later on to derive our main results.

In the reminder of the paper we focus our analysis on the distributions belonging to the exponential family. In particular, we consider density functions characterized by the following parametric expression:

$$f(x;\theta) = \exp\left\{\theta b(x) - A(\theta)\right\}, \tag{3.1}$$

where $\theta \in \Theta$ is a single real valued natural parameter and $A(\theta)$ is the cumulant generating function (or log normalizer). Since $\int_\Omega f(x;\theta)d\mu(x) = 1$, it is clear that $A(\theta)$ can be written as

$$A(\theta) = \log \int_\Omega \exp\left\{\theta b(x)\right\} d\mu(x). \tag{3.2}$$

Throughout the course of the discussion the true parameter will be denoted by θ_0. Next, we explore consistency, which is a basic requirement for a good estimator.

3.1. Consistency

Consider a monotone sequence of distortion parameters $\{q_n\}_{n\geq 1}$ such that $q_n \longrightarrow 1$ and impose the following requirements:

A.1 The parameter space Θ is compact.

A.2 For any $n \geq 1$, assume that $\sum_{i=1}^n \frac{\partial}{\partial\theta} L_{q_n} f\left(X_i;\theta\right)$ is continuous in θ.

A.3 Assume that $A^{(k)}(\theta_0 + k\theta(1-q_n)) < \infty$ $(k = 0,1,2)$ for any $\theta \in \Theta$, $n \geq 1$.

In similar contexts, the compactness condition is often used for technical reasons (see e.g., Wang et *Al.* [17]) as it is the case here.

Theorem 3.1. *Assume that θ_0 is an interior point in Θ. Under assumptions A.1 through A.3, for any sequence of MLq estimators of θ_0*

$$\widetilde{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} L_{q_n}\left[f(X_i;\theta)\right] \tag{3.3}$$

we have that $\widetilde{\theta}_n \xrightarrow{P} \theta_0$.

Remark 3.2. With probability converging to 1, the maximizer of $\sum_{i=1}^{n} L_{q_n}\left(f(X_i;\theta)\right)$ uniquely exits.

Although for a fixed $q \neq 1$, the MLqE is clearly asymptotically biased, a clear improvement is obtained by letting the distortion parameter depend on the sample size. If the degree of distortion diminishes as the amount of information carried by the sample increases, the new estimator gains the desirable consistency property.

As defined in 3.3, the MLqE is a consistent estimator of ξ. However, in the rest of the paper we shall discuss the reduction in terms of variance achieved by considering a slightly different target parameter. In particular, we consider θ_n^*, the value such that that $E_{\theta_0}\frac{\partial}{\partial\theta}L_{q_n}f\left(X;\theta\right)\Big|_{\theta=\theta_n^*} = 0$. In particular, θ_n^* can arbitrarily close to the true parameter θ_0 depending on the value of the distortion parameter q_n and $\theta_n^* = \theta_0$ when $q = 1$.

3.2. Asymptotic Normality

To obtain the asymptotic normality of MLqE, we introduce some additional conditions:

A.4 For each $\theta \in \Theta$ and $n \geq 1$, $\frac{\partial}{\partial\theta}L_{q_n}f\left(x;\theta\right)$ is twice continuously differentiable for every x.

A.5 Assume $A^{(k)}(\theta_0+k(1-q_n)\theta) < \infty$, $k = 3, 4$ for every θ in a neighborhood of θ_n^*.

Theorem 3.3. *If assumptions A.1 through A.5 are satisfied, then we have that*

$$\sqrt{n}\frac{\left(\widetilde{\theta}_n - \theta_n^*\right)}{\sigma_n} \xrightarrow{\mathcal{D}} N(0,1) \quad as \quad n \to \infty, \tag{3.4}$$

where

$$\sigma_n^2 = \frac{E_{\theta_0}\left[\frac{\partial}{\partial\theta}L_{q_n}f(X;\theta)\Big|_{\theta=\theta_n^*}\right]^2}{\left(E_{\theta_0}\frac{\partial^2}{\partial\theta^2}L_{q_n}f(X;\theta)\Big|_{\theta=\theta_n^*}\right)^2}. \tag{3.5}$$

Remark 3.4. When q_n converges to 1 slowly enough so that $\theta_n^* - \theta_0$ is of higher order than n^{-1}, then clearly $\sqrt{n}(\widetilde{\theta}_n - \theta_0)$ does not converge in distribution. Such a seemingly undesirable choice of q_n can be actually advantageous for estimating tail probability as will be seen.

The proof of the theorem is carried out by expanding (3.3) about θ_n^*. The regularity requirements are classical and concern mainly the smoothness of $\psi_n(x;\theta)$ and boundedness in probability of its second derivative about the value θ_n^*.

Example 3.1 (Continued). Consider $X_1, ..., X_n$ i.i.d. observations from an exponential distribution with true parameter λ_0. For any n, it can be easily found that $\lambda_n^* = \lambda_0/q_n$ is the unique zero of the equation $E_{\lambda_0}\left[\frac{\partial}{\partial\lambda}L_{q_n}f(X;\lambda)\right] = 0$ (see appendix B). The cumulant-generating function is

$$A(\lambda) = \log\int_0^\infty e^{-\lambda x}dx = \log\frac{1}{\lambda} \tag{3.6}$$

and differentiating k times gives

$$\frac{\partial^k}{\partial\lambda^k}A(\lambda) = \frac{(k-1)!(-1)^k}{\lambda^k}. \tag{3.7}$$

Thus, the conditions for asymptotic normality are met and the calculation in appendix B shows that the asymptotic variance for the MLqE in this case is

$$\sigma_n^2 = \frac{E_{\lambda_0}\left[\frac{\partial}{\partial\lambda} L_{q_n} f(X;\lambda)\Big|_{\lambda=\lambda_n^*}\right]^2}{\left(E_{\lambda_0} \frac{\partial^2}{\partial\lambda^2} L_{q_n}(X;\lambda)\Big|_{\lambda=\lambda_n^*}\right)^2} = \left(\frac{\lambda_0}{q_n}\right)^2 \left[\frac{q_n^2 - 2q_n + 2}{q_n^3 (2-q_n)^3}\right] \to \lambda_0^2. \quad (3.8)$$

as $n \to \infty$. By theorem 3.3, we conclude that

$$\left(\frac{n q_n^5 (2-q_n)^3}{q_n^2 - 2q_n + 2}\right)^{1/2} \left(\tilde{\lambda}_n - \frac{\lambda_0}{q_n}\right) \xrightarrow{\mathcal{D}} N\left(0, {\lambda_0}^2\right) \quad \text{as } n \to \infty. \quad (3.9)$$

Note that when $q_n = 1$ for any $n \geq 1$, the usual MLE estimator $\widehat{\lambda}_n$ is recovered and $\sqrt{n}\left(\widehat{\lambda}_n - \lambda_0\right) \xrightarrow{\mathcal{D}} N(0, {\lambda_0}^2)$ as $n \to \infty$.

4. Estimation of the Tail Probability

In this section we address the problem of tail probability estimation, using the popular "plug-in procedure, where the point estimate of the unknown parameter is substituted into the distribution of interest. We introduce the plug-in estimator for the tail probability, based on the MLq method and derive its asymptotic distribution. Moreover, in a suitable framework, we show that when the distortion parameter q is chosen according to the sample size, the ratio of the asymptotic Mean Squared Error of MLqE over that of MLE converges to 0.

Let $\alpha(x;\theta) = P_\theta(X \leq x)$ or $\alpha(x;\theta) = 1 - P_\theta(X \leq x)$, depending on whether we are considering the lower tail or the upper tail of the distribution. Suppose to observe $X_1, ..., X_n$, an i.i.d. sample from the true distribution $f(x;\theta_0)$, and let θ_n^* be defined as in the previous section. Under some conditions on $\alpha(x;\theta)$ we can expand $\alpha(x;\tilde{\theta}_n)$ about θ_n^*, obtaining:

$$\alpha(x_L;\widetilde{\theta_{0n}}) - \alpha(x_L;\theta_n^*) \simeq \alpha'(x_L;\theta_n^*)(\tilde{\theta}_n - \theta_n^*) + \frac{1}{2}\alpha''(x_L;\theta_n^*)(\tilde{\theta}_n - \theta_n^*)^2. \quad (4.1)$$

By theorem 3.1, combined with Slutsky's lemma, the asymptotic distribution of $\alpha(x_L;\widetilde{\theta}_n)$ can be found to be

$$\sqrt{n}\frac{\alpha(x_L;\widetilde{\theta}_n)-\alpha(x_L;\theta_n^*)}{\sigma_n^{1/2}\alpha'(x_L;\theta_n^*)} \xrightarrow{\mathcal{D}} N\left(0,1\right). \tag{4.2}$$

Note that when $q=1$, the usual maximum likelihood estimator $\widehat{\theta}_n$ is recovered and θ_n^* can be replaced by the true parameter θ_0 in the expression above.

However, it is known that in most applications a large amount of information is usually demanded in order to obtain accurate estimates of a small tail probability. As a consequence, the classical problem setup presented so far may be inadequate, as it ignores the relationship between x_L and the sample size n. We consider instead a framework that better reflects the nature of the problem.

4.1. Asymptotic normality of the Plug-in MLq estimator

Let $X_1,...,X_n$ be an i.i.d. sample from the true distribution $f(x;\theta_0)$. We are interested in estimating $\alpha(x_{L,n};\theta_0)$, where $x_{L,n}$ is a monotone sequence. Moreover, let $b(\cdot)$ be defined as in the functional form of the exponential family density introduced in (3.1). The next theorem extends the normality to the plug-in estimator $\alpha(x_{L,n};\widetilde{\theta}_n)$.

Theorem 4.1. *Let θ_n^* be as defined in the previous section and q_n a sequence converging to 1 as $n\rightarrow\infty$. Moreover, define*

$$h(x_{L,n};\theta_1) := \sup_{\theta\in[\theta_1-\delta,\theta_1+\delta]}\frac{\alpha''(x_{L,n};\theta)}{\alpha''(x_{L,n};\theta_1)},\ \delta>0. \tag{4.3}$$

Then, under assumptions A.1 through A.6, for every sequence $\{x_{L,n}\}_{n\geq1}$ such that $b(x_{L,n})h(x_{L,n};\theta_n^*)=o\left(n^{1/2}\right)$, we have

$$\sqrt{n}\frac{\left(\alpha(x_{L,n};\widetilde{\theta}_n)-\alpha(x_{L,n};\theta_n^*)\right)}{\sigma_n\alpha'(x_{L,n};\theta_n^*)} \xrightarrow{\mathcal{D}} N\left(0,1\right),$$

where σ_n is the asymptotic variance of $\widetilde{\theta}_n$ as in eq.(3.5).

Remark 4.2. The main requirement of the theorem on the order of the sequence $x_{L,n}$ it's easiest to be verified on a case by case basis. For instance, the tail probability of the exponential distribution in eq.(2.13) is $\alpha(x;\lambda) = e^{-x\lambda}$. Thus, $\alpha''(x;\lambda) = e^{-x\lambda}x^2$. In this case we have that $b(x_{L,n}) = -x_{L,n}$; moreover, given $\delta > 0$, one can see that $h(x_{L,n};\lambda_n^*) = e^{\delta x_{L,n}}$. Therefore, the main condition of the theorem reads:

$$b(x_{L,n})h(x_{L,n};\lambda_n^*) = e^{\delta x_{L,n}+\log x_{L,n}} = o(n^{1/2}), \tag{4.4}$$

which can be easily satisfied by choosing $x_{L,n} = o(\log n)$.

Given the size of the tail probability to be estimated, in many applications the quantity of interest is the corresponding quantile. In our setting, the quantile function is defined as $\rho(s;\theta) = \alpha^{-1}(s;\theta)$, $0 \le s \le 1$ and $\theta \in \Theta$. Next, we present the analogue of Theorem 4.1 for the plug-in estimate of the quantile.

Theorem 4.3. *Let $0 < s_n < 1$ be a nonincreasing sequence such that $s_n \searrow 0$ and let θ_n^* and q_n be defined as in theorem 4.1. Moreover, let*

$$h_1(s_n;\theta_1) := \sup_{\theta\in[\theta_1-\delta,\theta_1+\delta]} \frac{\rho''(s_n;\theta)}{\rho''(s_n;\theta_1)}, \tag{4.5}$$

and

$$h_2(s_n;\theta_1) := \sup_{\theta\in[\theta_1-\delta,\theta_1+\delta]} \frac{\rho''(s_n;\theta)}{\rho'(s_n;\theta_1)}, \tag{4.6}$$

where $\delta > 0$. Then, under assumptions A.1 through A.6, for every sequence s_n such that (i) $b\left(\rho(s_n;\theta_n^*)\right) h_1(s_n;\theta_n^*) = o\left(n^{1/2}\right)$ and the ratio $\rho''(s_n;\theta)/\rho'(s_n;\theta)$ is an indeterminate form $0/0$ or ∞/∞, or (ii) $h_2(s_n;\theta_n^*) = o\left(n^{1/2}\right)$, we have that

$$\sqrt{n}\frac{\left(\rho(s_n;\widetilde{\theta}_n) - \rho(s_n;\theta_n^*)\right)}{\sigma_n \rho'(s_n;\theta_n^*)} \xrightarrow{\mathcal{D}} N(0,1),$$

where σ_n is the asymptotic variance of $\widetilde{\theta}_n$ as in eq.(3.5).

In the new setting, we relate explicitly the amount of information available in the sample to both the "rarity" of the event under exam and the distortion parameter. In the next section we use this new framework to compare the proposed estimator of the tail probability based on MLqE, $\alpha(x_{L,n};\widetilde{\theta}_n)$, with the one based on the traditional MLE, $\alpha(x_{L,n};\widehat{\theta}_n)$.

4.2. Relative efficiency between MLE and MLqE

The results of asymptotic normality of the new estimator is crucial to compare the efficiency of the new estimator to that of the classical MLE. In particular, changing slightly the center of the asymptotic distribution can allow to gain a substantial variance reduction. Consider w_n and v_n, two estimators of a parametric function $g(\theta)$ that satisfy

$$\sqrt{n}\left(w_n - a_n\right)/\sigma_n \xrightarrow{\mathcal{D}} N\left(0,1\right), \tag{4.7}$$

$$\sqrt{n}\left(v_n - b_n\right)/\tau_n \xrightarrow{\mathcal{D}} N\left(0,1\right), \tag{4.8}$$

for some sequences a_n, b_n, $\sigma_n > 0$ and $\tau_n > 0$. The *bias adjusted relative efficiency* of w_n with respect to v_n as

$$\Lambda(w_n, v_n) := \frac{\left(b_n - g(\theta)\right)^2 + \tau_n^2/n}{\left(a_n - g(\theta)\right)^2 + \sigma_n^2/n}. \tag{4.9}$$

In general the relative efficiency between MLE and MLqE is best to be evaluated on a case-by-case basis. In the following example we continue to discuss the transparent, yet important case of the exponential distribution.

Example 4.1 (Continued). Consider a sample $X_1, ..., X_n$ from the exponential distribution in (3.1). In this case, we have $\alpha(x_{L,n};\lambda) = e^{-\lambda x_{L,n}}$ and $\alpha'(x_{L,n};\lambda) = -x_{L,n}e^{-\lambda x_{L,n}}$. For properly chosen sequences $x_{L,n}$ and q_n we have that

$$\sqrt{n}\frac{\left(e^{-\widetilde{\lambda}_n x_{L,n}} - e^{-\frac{\lambda_0}{q_n}x_{L,n}}\right)}{\sigma_n x_{L,n} e^{-\frac{\lambda_0}{q_n}x_{L,n}}} \xrightarrow{\mathcal{D}} N\left(0,1\right), \tag{4.10}$$

where the asymptotic variance of $\widetilde{\lambda}_n$ is computed in Appendix B as

$$\sigma_n = \left(\frac{\lambda_0}{q_n}\right)^2 \left[\frac{q^2 - 2q_n + 2}{q_n^3 (2-q_n)^3}\right]. \tag{4.11}$$

When $q_n = 1$ for any $n \geq 1$, we recover the usual plug-in estimator based traditional MLE, $\widehat{\lambda}_n$,

$$\sqrt{n}\frac{\left(e^{-\widetilde{\lambda}_n x_{L,n}} - e^{-\lambda_0 x_{L,n}}\right)}{\lambda_0 x_{L,n}^2 e^{-2\lambda_0 x_{L,n}}} \xrightarrow{\mathcal{D}} N(0,1). \tag{4.12}$$

Let $k_n := \dfrac{q_n^2 - 2q_n + 2}{q_n^3 (2-q_n)^3}$. The relative efficiency of $\alpha(x_L; \widehat{\lambda}_n)$ compared to $\alpha(x_L; \widetilde{\lambda}_n)$ is

$$\Lambda(\alpha(x_L;\widehat{\lambda}_n), \alpha(x_L;\widetilde{\lambda}_n)) \tag{4.13}$$

$$= \frac{\left(e^{-\frac{\lambda_0}{q_n}x_{L,n}} - e^{-\lambda_0 x_{L,n}}\right)^2 + n^{-1}\lambda_0^2 q_n^{-2} k_n x_{L,n}^2 e^{-2\frac{\lambda_0}{q_n}x_{L,n}}}{n^{-1}\lambda_0^2 x_{L,n}^2 e^{-2\lambda_0 x_{L,n}}} \tag{4.14}$$

$$= \frac{n}{\lambda_0^2 x_{L,n}^2}(e^{-x_{L,n}(\lambda_0/q_n - \lambda_0)} - 1)^2 + q_n^{-2} k_n e^{-2x_{L,n}(\lambda_0/q_n - \lambda_0)} \tag{4.15}$$

and by adding and subtracting $q_n^{-2}k_n$, we obtain

$$\frac{n\left(1-\frac{1}{q_n}\right)^2}{\lambda_0^2 x_{L,n}^2}(e^{-x_{L,n}(\lambda_0/q_n-\lambda_0)} - 1)^2 + q_n^{-2}k_n\left(e^{-2x_{L,n}(\lambda_0/q_n-\lambda_0)} - 1\right) + q_n^{-2}k_n. \tag{4.16}$$

Recall that $(1-r)L_r(u) = u^{1-r} - 1$, $r > 0$. Thus, the expression above can be written as

$$\frac{n\left(1-\frac{1}{q_n}\right)^2}{\lambda_0^2 x_{L,n}^2} L_{1/q_n}\left(e^{x_{L,n}\lambda_0}\right)^2 + q_n^{-2}k_n\left(1-\frac{1}{q_n}\right) L_{1/q_n}\left(e^{2x_{L,n}\lambda_0}\right) + q_n^{-2}k_n \tag{4.17}$$

$$< n\left(1-\frac{1}{q_n}\right)^2 + k_n^*\left(1-\frac{1}{q_n}\right)2x_{L,n}\lambda_0 + k_n^*, \tag{4.18}$$

where $k_n^* = q_n^{-2}k_n$ and the last inequality holds because $L_{1/q}(u) < \log(u)$ for any $u > 0$ and $q < 1$. Next, we impose eq.(4.18) to be smaller than 1 and solve for $x_{L,n}$, obtaining

$$x_{L,n} > \frac{n\left(1-\frac{1}{q_n}\right)^2 + k_n^* - 1}{2\left(\frac{1}{q_n}-1\right)k_n^*\lambda_0} > \frac{n\left(1-\frac{1}{q_n}\right)^2}{2\left(\frac{1}{q_n}-1\right)k_n^*\lambda_0}, \tag{4.19}$$

Since, q_n is a monotone sequence, one can easily verify that k_n^* is nonincreasing in n and $k_n^* \geq 1$ for any $n \geq 1$. Therefore, from the inequality (4.19) we can derive the following lower bound for q_n

$$q_n > \left(\frac{2\lambda_0 x_{L,n} k_1^*}{n} + 1\right)^{-1} := T_n. \tag{4.20}$$

Hence, in order to guarantee $\Lambda < 1$ it sufficies to choose a sequence q_n such that $T_n < q_n < 1$, when the sample size is large enough. Note that the bound depends on the size of the probability to be estimated through $x_{L,n}$. This simple calculation provides useful insights on the choice of the sequence q_n in accordance to the size of the tail probability to be estimated. If q_n approaches too quickly 1, the gain obtained in terms of variance vanishes rapidly when n becomes large. Conversely, if q_n converges to 1 too slowly the bias part dominates the variance and the MLE outperforms the MLqE.

5. Monte Carlo Simulations

In this section, we illustrate the properties of the proposed estimator through Monte Carlo simulations on a simple univariate problem. Specifically, the performance of the MLqE is evaluated for the case of the exponential distribution in eq.(3.1). The aim of the simulation study is: (i) to explore the performance of the MLq estimator in finite sample situations in terms of mean squared error; and (ii) to inquire the reliability of confidence intervals constructed via asymptotics and bootstrap methods. The standard MLE estimator is used as a benchmark throughout the study.

5.1. Mean Squared Error: role of the distortion parameter q

In the first group of simulations, we compare the estimators of the true tail probability $\alpha = \alpha(x_L; \lambda_0)$, obtained via the ML$q$ method and the traditional maximum likelihood approach. Particularly, we are interested in assessing the relative performance of the two estimators for different choices of the sample size by taking the ratio between the two mean squared errors, $MSE(\hat{\alpha}_n)/MSE(\tilde{\alpha}_n)$. The simulations are structured as follows:

(i) For any given sample size $n \geq 2$, a number $B = 10000$ of Monte Carlo samples $X_1, ..., X_n$ is generated from an exponential distribution with parameter $\lambda_0 = 1$.

(ii) For each sample, the MLq and ML estimates of α, respectively $\tilde{\alpha}_{n,k} = \alpha(x_L; \tilde{\lambda}_{n,k})$ and $\hat{\alpha}_{n,k} = \alpha(x_L; \hat{\lambda}_{n,k})$, $k = 1, ..., B$, are obtained. The MLE of λ is simply computed as $\hat{\lambda}_{n,k} = n(\sum_{i=1}^n X_i)^{-1}$, whereas the ML$q$ estimator, $\tilde{\lambda}_n$, is computed by solving numerically the Lq-Likelihood equation (2.1). The optimization is performed by using variable metric algorithm (e.g., see Broyden [5]), where $\hat{\lambda}_{n,k}$ is chosen as starting value.

(iii) For each sample size n, the relative performance between the two estimators is evaluated by the ratio

$$\widehat{R}_n = \frac{MSE_{MC}(\hat{\alpha}_n)}{MSE_{MC}(\tilde{\alpha}_n)} = \frac{\sum_{k=1}^B (\hat{\alpha}_{n,k} - \alpha)^2}{\sum_{k=1}^B (\tilde{\alpha}_{n,k} - \alpha)^2},$$

where MSE_{MC} denotes the Monte Carlo estimates of the mean squared error. In addition, let $\overline{y}_1 = B^{-1} \sum_{k=1}^B (\hat{\alpha}_{n,k} - \alpha)^2$ and $\overline{y}_2 = B^{-1} \sum_{k=1}^B (\tilde{\alpha}_{n,k} - \alpha)^2$. By the central limt theorem, for large values of B we have that $\overline{\mathbf{y}} = (\overline{y}_1, \overline{y}_2)'$ converges weakly to a bivariate normal distribution with mean $\mu = (MSE(\hat{\alpha}_n), MSE(\tilde{\alpha}_n))'$ and covariance matrix Σ. Thus, the standard error for $\widehat{R}_n$ can be computed via the multivariate Delta

Method [3] as

$$se\left(\widehat{R}_n\right) = B^{1/2}\left(\frac{\widehat{\sigma}_{11}}{\overline{y}_2} - 2\widehat{\sigma}_{12}\frac{\overline{y}_1}{\overline{y}_2^3} + \widehat{\sigma}_{22}\frac{\overline{y}_1^2}{\overline{y}_2^4}\right)^{1/2}$$

where $\widehat{\sigma}_{11}$, $\widehat{\sigma}_{22}$ and $\widehat{\sigma}_{12}$ denote respectively the Monte Carlo estimates for the components of the covariance matrix Σ.

The procedure described above is repeated for different choices of the sample size, under three different experimental scenarios concerning the tail probability α and the distortion parameter q.

Case 1: fixed α and q. Figure 5.1 illustrates the behavior of $\widehat{R}_n$ for several choices of the sample size. In general, we observe that for relatively small sample sizes, $\widehat{R}_n > 1$ and the MLqE clearly outperforms the traditional MLE. Such a behavior is much more accentuated for smaller values of the tail probability to be estimated. In contrast, when the sample size is larger, the bias component plays an increasingly relevant role and eventually we observe that $\widehat{R}_n < 1$. This case is presented in figure 5.1/(a) for values of the true tail probability $\alpha = .01, .005, .003$ and a fixed distortion parameter $q = 0.5$. Moreover, the results presented in figure 5.1/(b) show that smaller values of the distortion parameter q accentuate the benefits attainable in small sample situation.

Case 2: fixed α and $q_n \nearrow 1$. In the second experimental setting, illustrated in figure 1, the tail probability α is fixed, while we let q_n be a sequence such that $q_n \nearrow 1$ and $0 < q_n < 1$. For illustrative purposes we choose the sequence $q_n = \left[1/2 + e^{0.3(n-20)}\right] / \left[1 + e^{0.3(n-20)}\right]$, $n \geq 2$ and study R_n for different choices of the true tail probability to be estimated. For small values of the sample size, the chosen sequence q_n converges relatively slowly to 1 and the distortion parameter produces benefits in terms of variance. In contrast, when the sample size becomes larger, q_n adjusts quickly to one. As a consequence, the MLqE gains for large samples the same behavior shown by the traditional MLE.

Case 3: $\alpha_n \searrow 0$ *and* $q_n \nearrow 1$. The last experimental setting of this sections treats the case where both the true tail probability and the distortion parameter change depending on the sample size. In particular, the values are following the theoretical results considerations from theorem ??. We consider sequences of distortion parameters converging slowly relative to the sequence of quantiles $x_{L,n}$. In particular we set $q_n = 1 - [10\log(n+10)]^{-1}$ and $x_{L,n} = n^{\frac{1}{2+\delta}}$. In the simulation described in figure 3 we illustrate the behavior of the estimator for $\delta = 0.5, 1.0$ and 1.5, confirming the theoretical findings discussed in 4.

5.2. *Asymptotic and Bootstrap Confidence Intervals*

The main objective of the simulations presented in this section is twofold: (a) to study the reliability of MLqE based confidence intervals constructed using three commonly used methods: asymptotics, parametric and nonparametric bootstraps; (b) to compare the results with those obtained using MLE.

When making inferences using the MLqE, an important issue is the selection of the distortion parameter q, according to the size of the sample under exam. One possible manner to handle this problem is to minimize the asymptotic mean squared error of the estimator. In the case of the exponential distribution, by theorem 4.3 we know the following expression for the asymptotic mean squared error:

$$MSE(q,\lambda_0) = \left(e^{-\frac{\lambda_0}{q}u} - e^{-\lambda_0 u}\right)^2 + \left(\frac{\lambda_0}{q}\right)^2 n^{-1} \left(\frac{q^2-2q+2}{q^3(2-q)^3}\right) u^2 e^{-2\frac{\lambda_0}{q}u}. \tag{5.1}$$

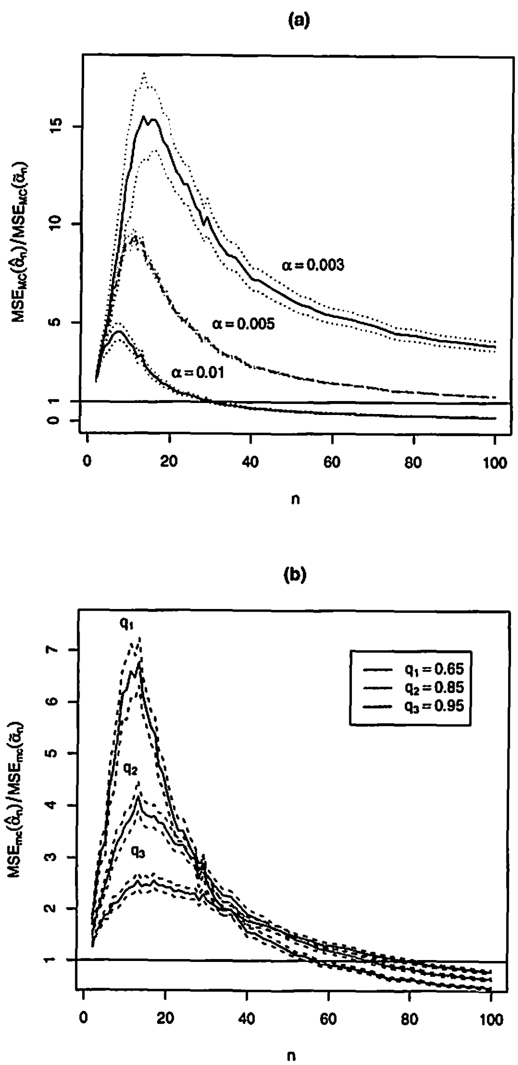

FIG 1. *Monte Carlo Mean Squared Error ratio computed from $B = 10000$ samples of size n. In (a) we use a fixed distortion parameter $q = 0.5$ and true tail probability $\alpha = 0.01, 0.005, 0.003$. The dashed lines represent 99% confidence bands. In (b) we set $\alpha = 0.003$ and the distortion parameters $q = 0.65, 0.85, 0.95$. The dashed lines represent 90% confidence bands.*

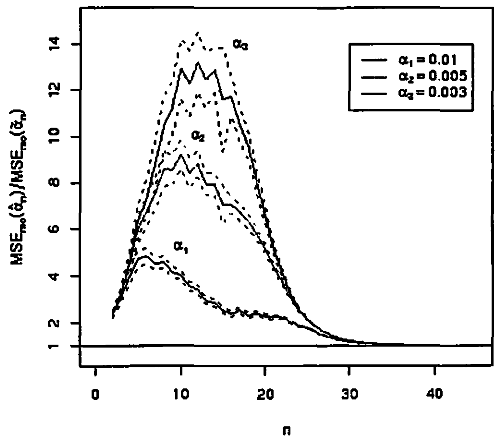

FIG 2. *Monte Carlo Mean Squared Error ratio computed from* $B = 10000$ *samples of size* n*, for different values of the true probability (*$\alpha = .01, .005, .003$*). The distortion parameter is computed as* $q_n = \left[1/2 + e^{0.3(n-20)}\right] / \left[1 + e^{0.3(n-20)}\right]$*. The dashed lines represent 99% confidence bands.*

However, since λ_0 is unknown, we consider

$$q^* = \underset{q \in (0,1)}{\arg\min} \left\{ MSE(q, \widehat{\lambda}) \right\}, \tag{5.2}$$

where $\widehat{\lambda}$ is an asymptotically unbiased estimator of λ_0. For the simulations presented in this section we use the MLE. The structure of simulations is similar to that of section 5.1:

(i) For any given sample size n, 1000 Monte Carlo samples $X_1, ..., X_n$ are generated from an exponential distribution with parameter $\lambda_0 = 1$.

(ii) For each sample, first we compute $\widehat{\lambda}_n$, the MLE of λ_0; we substitue $\widehat{\lambda}_n$ in eq. (5.1) and solve it numerically in order to obtain q^*.

(iii) For each sample, the MLq and ML estimates of the tail probability α are obtained. The standard errors of the estimates are computed using three different methods: the asymptotic formula derived in (4.10),

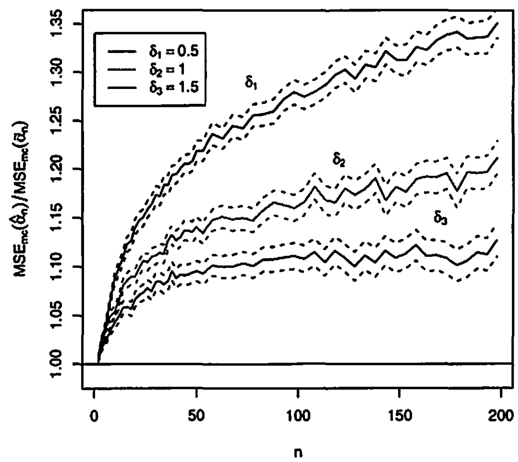

FIG 3. *Monte Carlo Mean Squared Error ratio computed from* $B = 10000$ *samples of size* n. *We use sequences* $q_n = 1 - [10\log(n+10)]^{-1}$ *and* $x_{L,n} = n^{\frac{1}{2+\delta}}$ ($\delta = 0.5, 1.0$ *and* 1.5). *The dashed lines represent 99% confidence bands.*

nonparametric and parametric bootstrap. The number of replicates emplyed in bootstrap resampling is 500. We also construct 95% confidence intervals and check the coverage of the true value α.

In table 1 we show the Monte Carlo estimates of the following quantities for different choices of n: $\widehat{\alpha}_n$ and $\widetilde{\alpha}_n$, their standard deviations and the standard errors computed with the three methods described above. In addition, we report the Monte Carlo average of the optimal distortion parameters q^*. Note that when $q^* = 1$, the results refer to the MLE case.

First note that, not surprisingly, q^* approaches 1 as the sample size increases. As a consequence, when the sample size is small, the MLqE pays an higher cost in terms of bias, compensated by a smaller standard deviation. Conversely, when n is larger and the bias becomes increasingly relevant, the MLqE trades bias for variance. As far as the standard errors are concerned, the asymptotic method and the parametric bootstrap seem to provide the values closer to the Monte Carlo standard deviation for all the considered sample sizes.

In table 2 we compare the accuracy of 95% confidence intervals constructed using the estimates of coverage probabilities. In addition, we compute the relative length of the intervals for MLE over those for MLqE. Although the coverage probability for MLqE is slightly smaller than that of MLE (in the order of 1%), we observe a substantial reduction in the interval length for all the considered cases. The most evident benefits occur when the sample size is small. Furthermore, note that in general the intervals computed via parameteric bootstrap outperform the other two methods in terms of coverage and length.

TABLE 1

MC estimates and standard deviations of α for different sample sizes, along with the MC estimates of the standard error computed using: (i) asymptotics, (ii) boostrap and (iii) parametric bootstrap. For the MLqE we also report the MC mean of the optimal parameters q^* ($q = 1$ corresponds to the MLE). The true tail probability is $\alpha = .01$.

n	q^*	Estimate	St.Dev.	se_{asy}	se_{boot}	se_{pboot}
15	.939	.009489	.010975	.010472	.011923	.010241
	1.000	.013464	.014830	.013313	.013672	.015090
25	.959	.009693	.008417	.008470	.009134	.008298
	1.000	.012108	.010517	.009919	.010227	.010950
50	.977	.010108	.006261	.006326	.006575	.006249
	1.000	.011385	.007354	.006894	.007083	.007318
100	.988	.010158	.004480	.004568	.004680	.004549
	1.000	.010789	.004908	.004778	.004880	.004943
500	.998	.010006	.002014	.002052	.002061	.002050
	1.000	.010122	.002055	.002070	.002073	.002087

TABLE 2

MC coverage rate of 95% confidence intervals for α, computed using (i) asymptotics, (ii) boostrap and (iii) parametric bootstrap. For the MLqE we report the MC mean of the optimal parameters q^* ($q = 1$ corresponds to the MLE). RL is the relative lenght of the intervals. The true tail probability is $\alpha = .01$. ($\alpha = .01$).

		Asympt.		Boot.		Par.Boot.	
n	q^*	Coverage(%)	RL	Coverage(%)	RL	Coverage(%)	RL
15	.939	79.2	1.271	84.8	1.147	87.4	1.473
	1.000	80.9		84.6		88.1	
25	.958	83.4	1.171	86.2	1.120	88.2	1.320
	1.000	84.3		87.2		89.6	
50	.977	87.1	1.089	88.9	1.077	89.3	1.171
	1.000	88.4		89.4		89.9	
100	.988	91.1	1.046	91.9	1.043	92.0	1.087
	1.000	92.2		92.4		93.0	
500	.998	94.5	1.009	94.0	1.006	94.1	1.018
	1.000	94.7		94.5		94.7	

6. Concluding Remarks

In this work we have introduced the MLqE, a new estimator of the parameters inspired by a class of generalized information measures. The new estimator appears to be a natural extension of the classical MLE: it preserves the large sample properties of the MLE, while it gains the control over a distortion parameter q, allowing to modify the trade-off between bias in small or moderate sample situations. Although we have considered the MLqE for the specific purpose of tail probability estimation, this research is intented to be a starting point. Its variance reduction capabilities can be possibly studied and exploited in other inferential settings, including conditional probability estimation and generalized linear models with covariates.

We emphasized both from theoretical and empirical standpoints the meaning of the link between q and the sample size. The asymptotic rate of convergence of the mean squared error of the MLqE over that of MLE provides insights about the behvior q with respect the sample size and tail probability (or quantile) to be estimated. In addition, the Monte Carlo simulations support the theoretical findings showing that when the sample size is small or moderate relative to the tail probability to be estimated and $q < 1$, the MLqE successfully trades bias for variance, obtaining an overall reduction of the mean squared error. However, the practical choice of the parameter q, especially for small and moderate sample size situations is beyond the purpose of this work and requires further scrutiny.

Appendix A - Proofs

In all of the following proofs we denote $\psi_n(\theta) := n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\theta} L_{q_n}\left(f(X_i;\theta)\right)$. Since $f(x;\theta) = e^{\theta b(x)-A(\theta)}$, we can write

$$\psi_n(\theta) = \frac{1}{n}\sum_{i=1}^n e^{(1-q_n)(\theta b(X_i)-A(\theta))}\left(b(X_i) - A'(\theta)\right). \tag{6.1}$$

The MLqE is found by setting $\psi_n(\theta) = 0$ and solving for θ. Note that setting $q_n = 1$ in the above expression gives the usual maximum likelihood estimating equation $n^{-1}\sum_{i=1}^n \left(b(X_i) - A'(\theta)\right)$. In addition, define $\varphi(x,\theta) := \theta b(x) - A(\theta)$, so that the exponential family can be expressed as $f(x;\theta) = e^{\varphi(x,\theta)}$. When clear from the context, $\varphi(x,\theta)$ is denoted by φ.

Proof of Theorem 3.1

Define $\psi(\theta) := E_{\theta_0}\frac{\partial}{\partial\theta}\log\left(f(X;\theta)\right)$. Since f has the form in (3.1), we can write $\psi(\theta) = E_{\theta_0}\left(b(X_i) - A'(\theta)\right)$. First, we want to show that for all $\theta \in \Theta$,

$\psi_n(\theta) \xrightarrow{p} \psi(\theta)$, or equivalently

$$\frac{1}{n}\sum_{i=1}^{n}\left(e^{(1-q_n)(\theta b(X_i)-A(\theta))}-1\right)(b(X_i)-A'(\theta)) \xrightarrow{p} 0. \tag{6.2}$$

Let $S(X_i;\theta) := e^{(1-q_n)(\theta b(X_i)-A(\theta))}-1$ and $T(X_i;\theta) := b(X_i)-A'(\theta)$. By the Cauchy-Schwartz inequality, we can bound the left hand side of eq. (6.2) as follows

$$|LHS| \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}S(X_i;\theta)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}T(X_i;\theta)^2}. \tag{6.3}$$

Note that

$$E_{\theta_0}T(X;\theta)^2 = E_{\theta_0}(b(X)-A'(\theta_0)+A'(\theta_0)-A'(\theta))^2 = A''(\theta_0)+(A'(\theta_0)-A'(\theta))^2 < \infty, \tag{6.4}$$

and

$$E_{\theta_0}S(X_i;\theta)^2 = E_{\theta_0}e^{2(1-q_n)(\theta b(X)-A(\theta))} - 2E_{\theta_0}e^{(1-q_n)(\theta b(X)-A(\theta))} + 1. \tag{6.5}$$

Observe that for each θ,

$$\begin{aligned} E_{\theta_0}e^{2(1-q_n)(\theta b(X)-A(\theta))} =& \int_\Omega e^{(\theta_0+2(1-q_n)\theta b(X)-A(\theta_0)+2(1-q_n)\theta b(X)}d\mu(x)e^{-A(\theta_0)-2(1-q_n)A(\theta)} \\ =& e^{A(\theta_0+2(1-q_n)\theta)-A(\theta_0)-2(1-q_n)A(\theta)} \longrightarrow 1 \end{aligned}$$

and similarly, $E_{\theta_0}e^{(1-q_n)(\theta b(X)-A(\theta))} \to 1$ as $n\to\infty$. Consequently, $E_{\theta_0}S(X_i;\theta)^2 \to \infty$ and the right hand side in (6.3) converges in probability to zero as $n\to\infty$.

Next, let B_n be the event that $\psi_n(\theta)$ is non-decreasing in $\theta \in \Theta$. We want to show that $P(B_n)\to 0$ as $n\to\infty$. First, differentiate $\psi_n(\theta)$, obtaining

$$\frac{\partial}{\partial\theta}\psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}e^{(1-q_n)(\theta b(X_i)-A(\theta))}\left[(1-q_n)(b(X_i)-A'(\theta))^2 - A''(\theta)\right].$$

So, we have that $\frac{\partial}{\partial\theta}\psi_n(\theta) < 0$ when

$$A''(\theta)\frac{1}{n}\sum_{i=1}^{n}e^{(1-q_n)(\theta b(X_i)-A(\theta))} > (1-q_n)\frac{1}{n}\sum_{i=1}^{n}e^{(1-q_n)(\theta b(X_i)-A(\theta))}(b(X_i)-A'(\theta))^2. \tag{6.6}$$

Since Θ is compact, $\sup_{\theta\in\Theta} e^{-A(\theta)} \leq c_1 < \infty$ and $\sup_{\theta\in\Theta} A'(\theta)^2 \leq c_2 < \infty$ for some constants c_1 and c_2. Moroever,

$$\sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} e^{(1-q_n)(\theta b(X_i)-A(\theta))}\left(b(X_i)-A'(\theta)\right)^2 \tag{6.7}$$

$$\leq \frac{1}{n}\sum_{i=1}^{n} \sup_{\theta\in\Theta} e^{(1-q_n)(\theta b(X_i)-A(\theta))}\left(b(X_i)-A'(\theta)\right)^2 \tag{6.8}$$

$$\leq \frac{2}{n}\sum_{i=1}^{n} c_1\left(e^{(1-q_n)\theta_0 b(X_i)} + e^{(1-q_n)\theta_1 b(X_i)}\right)\left(b(X_i)^2 + c_2\right), \tag{6.9}$$

where θ_0 and θ_1 are the boundary points in Θ. Moreover, by assumption A.3 one can see that the expectation of (6.9) is some constant $K < \infty$. Therefore, given some small constant $t > 0$ by Markov's inequality, we have that $\inf_{\theta\in\Theta} A''(\theta) \geq c_3 > 0$ for some constant c_3

$$P\left((1-q_n)\frac{1}{n}\sum_{i=1}^{n} e^{(1-q_n)(\theta b(X_i)-A(\theta))}\left(b(X_i)-A'(\theta)\right)^2 > t\right) \leq t^{-1}(1-q_n)K, \tag{6.10}$$

for some constant K.

Next, consider the left hand side of the inequality (6.6). Since Θ is compact, $\inf_{\theta\in\Theta} e^{-A(\theta)} \geq c_3 > 0$ and $\sup_{\theta\in\Theta} A''(\theta) \geq c_4 > 0$, for some constants c_3 and c_4. Moreover, we have that

$$\frac{1}{n}\sum_{i=1}^{n} e^{(1-q_n)(\theta b(X_i)-A(\theta))} \geq \frac{c_4}{n}\sum_{i=1}^{n} e^{(1-q_n)\theta b(X_i)} \geq \frac{c_4}{n}\sum_{i=1}^{n} Y_{i,n}, \tag{6.11}$$

where $Y_{i,n} = \min\left\{e^{(1-q_n)\theta_0 b(X_i)}, e^{(1-q_n)\theta_1 b(X_i)}\right\}$ and θ_0, θ_1 are the boundary points in Θ. Therefore, by the law of large numbers, (6.11) is bounded from below in probability by a positive constant.

From (6.10) and (6.11) we have that the function $\psi_n(\theta)$ is nondecreasing with probability converging to 1 as $n \to \infty$. The desired result of consistency follows by applying lemma 5.10 p.47 [2].

Proof of Theorem 3.3

Let $\underline{X} := (X_1, ..., X_n)$ and define

$$\psi_n(\underline{X}; \theta) := n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} L_{q_n} f(X_i, \theta) = n^{-1} \sum_{i=1}^{n} e^{(1-q_n)\varphi(X_i,\theta)} \varphi'(X_i; \theta),$$

and

$$\psi(\underline{X}; \theta) := E_{\theta_0} \frac{\partial}{\partial \theta} L_{q_n} f(X; \theta) = E_{\theta_0} e^{(1-q_n)\varphi(X_i;\theta)} \varphi'(X; \theta).$$

By Taylor's theorem there exist a value $\widetilde{\theta}$, between θ_n^* and $\widetilde{\theta}_n$, such that with probability converging to one we have

$$\begin{aligned} 0 &= \psi_n\left(\underline{X}; \widetilde{\theta}_n\right) \\ &= \psi_n\left(\underline{X}; \theta_n^*\right) + \dot{\psi}_n\left(\underline{X}; \theta_n^*\right)\left(\widetilde{\theta}_n - \theta_n^*\right) + \frac{1}{2}\ddot{\psi}_n\left(\underline{X}; \widetilde{\theta}_n\right)\left(\widetilde{\theta}_n - \theta_n^*\right)^2, \end{aligned} \quad (6.12)$$

where $\dot{\psi}_n$ and $\ddot{\psi}_n$ denote first and second derivative with respect the parameter. We can rewrite the above expression as

$$0 = \sqrt{n}\frac{\psi_n\left(\underline{X}; \theta_n^*\right)}{\dot{\psi}\left(\theta_n^*\right)} + \frac{\dot{\psi}_n\left(\underline{X}; \theta_n^*\right)}{\dot{\psi}\left(\theta_n^*\right)}\sqrt{n}\left(\widetilde{\theta}_n - \theta_n^*\right) + \frac{1}{2}\frac{\ddot{\psi}_n\left(\underline{X}; \widetilde{\theta}\right)}{\dot{\psi}\left(\theta_n^*\right)}\sqrt{n}\left(\widetilde{\theta}_n - \theta_n^*\right)^2, \quad (6.13)$$

where $\dot{\psi}(X; \theta) = E_{\theta_0} \frac{\partial^2}{\partial \theta^2} L_{q_n} f(X; \theta)$. We take the following steps to derive asymptotic normality.

Step 1. We first show that the first term in (6.13) converges in distribution. Let $z_{n,i} := \frac{\partial}{\partial \theta} L_{q_n} f(X_i, \theta)\Big|_{\theta_n^*}$. Since $z_{n,i}$ $(1 \leq i \leq n)$ forms a triangular array where $z_{n,i}$ are rowwise i.i.d, the Lindberg-Feller condition must be satisfied (i.e., see Ferguson p. 27 [3]). In our case the condition reads: given $\varepsilon > 0$,

$$\left(E z_{n,1}^2\right)^{-1} E\left[z_{n,1}^2 I\left(|z_{n,1}| \geq \varepsilon\sqrt{n}\left(E z_{n,1}^2\right)^{1/2}\right)\right] \to 0 \text{ as } n \to \infty. \quad (6.14)$$

By Hölder's and Chebychev's inequalities we can write

$$\left(E z_{n,1}^2\right)^{-1} E\left[z_{n,1}^2 I\left(|z_{n,1}| \geq \varepsilon\sqrt{n}\left(E z_{n,i}^2\right)^{1/2}\right)\right] \leq \frac{1}{\varepsilon^{2/3} n^{1/3}}\left(E z_{n,1}^2\right)^{-1}\left(E\left[z_{n,1}^3\right]\right)^{2/3}.$$

Next, denote $\mu_{n,k} = \theta_0 + k(1-q_n)\theta_n^*$. One can see that

$$\begin{aligned}\left(E\left[z_{n,1}^3\right]\right)^{2/3} &= \left(E_{\mu_{n,3}}\left[b(X) - A'(\theta_n^*)\right]^3\right)^{2/3} \\ &\quad \times \exp\left\{-\frac{2}{3}A(\theta_0) - 2(1-q_n)A(\theta_n^*) + \frac{2}{3}A(\mu_{n,3})\right\} \\ &= \left(A'''(\mu_{n,3}) + (A'(\mu_{n,3}) - A'(\theta_n^*))^3 + 3A''(\mu_{n,3})A'(\theta_n^*)\right)^{2/3} \\ &\quad \times \exp\left\{-\frac{2}{3}A(\theta_0) - 2(1-q_n)A(\theta_n^*) + \frac{2}{3}A(\mu_{n,3})\right\}.\end{aligned}$$

Since the parameter space is assumed to be compact, the above quantity is upper bounded by some constant.

$$\begin{aligned}E\left[z_{n,1}^2\right] &= E_{\theta_0}\left[\exp\left\{2(1-q_n)\varphi(X,\theta_n^*)\right\}\varphi'(X,\theta_n^*)^2\right] \\ &= E_{\mu_{n,2}}\left[b(X) - A'(\theta_n^*)\right]^2 \\ &\quad \times \exp\left\{-A(\theta_0) - 2(1-q_n)A(\theta_n^*) + A(\mu_{n,2})\right\}.\end{aligned}$$

Note that the quadratic function $E_{\mu_{n,2}}\left[b(X) - a\right]^2$ is minimized when $a = A'(\mu_{n,2})$. Thus, we have that $E\left[z_{n,1}^2\right]$ is lower bounded away from zero. Therefore, the condition (6.14) holds, providing the desired result.

Step 2. Next, we want to show that $\dot{\psi}_n\left(\underline{X}, \theta_n^*\right)/\dot{\psi}\left(X, \theta_n^*\right) \xrightarrow{P} 1$. Given $\varepsilon > 0$, we have that

$$P_{\theta_0}\left(\left|\frac{\dot{\psi}_n\left(\underline{X}, \theta_n^*\right)}{\dot{\psi}\left(X, \theta_n^*\right)} - 1\right| > \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\frac{E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}L_{q_n}\left(f\left(X;\theta\right)\right)\Big|_{\theta_n^*}\right]^2}{\left(E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}L_{q_n}\left(f\left(X;\theta\right)\right)\Big|_{\theta_n^*}\right]\right)^2} \tag{6.15}$$

by the i.i.d. assumption and Chebychev's inequality. First consider the numerator in (6.15) and see that

$$\begin{aligned}&E_{\theta_0}\left[\exp\left\{2(1-q_n)\varphi\right\}\left((\varphi')^2(1-q_n) + \varphi''\right)^2\right] \\ &\leq E_{\theta_0}\left[\exp\left\{2(1-q_n)\varphi\right\}\left((\varphi')^2 + \varphi''\right)^2\right] \\ &\leq 2E_{\mu_{n,2}}\left[\left(b(X) - A'(\theta_n^*)\right)^4 + A''(\theta_n^*)^2\right] \\ &\times \exp\left\{-A(\theta_0) - 2(1-q_n)A(\theta_n^*) + A(\mu_{n,2})\right\}\end{aligned}$$

where the last passage follows from the basic inequality $(a+b)^2 \leq 2a^2+2b^2$ (for any $a,b \in \mathcal{R}$). Further, note that

$$
\begin{aligned}
&E_{\mu_{n,2}}\left[(b(X)-A'(\theta_n^*))^4 + A''(\theta_n^*)^2\right] \\
&= E_{\mu_{n,2}}\left[(b(X)-A'(\mu_{n,2})+A'(\mu_{n,2})-A'(\theta_n^*))^4 + A''(\theta_n^*))^2\right] \\
&= A^{(4)}(\mu_{n,2}) + (A'(\mu_{n,2})-A'(\theta_n^*))^4 + A''(\theta_n^*))^2 \\
&+ 4A^{(3)}(\mu_{n,2})(A'(\mu_{n,2}-A'(\theta_n^*)) + 6A''(\mu_{n,2})(A'(\mu_{n,2})-A'(\theta_n^*))^2.
\end{aligned}
$$

Next, from the denominator in (6.15), write

$$
\begin{aligned}
&\left(E_{\theta_0}\left[\exp\left\{(1-q_n)\varphi\right\}\left((\varphi')^2(1-q_n)+\varphi''\right)\right]\right)^2 \\
&= \left(E_{\mu_{n,1}}\left[b(X)-A'(\theta_n^*)\right]^2(1-q_n) - A(\theta_n^*)''\right)^2 \\
&\times \exp\left\{-2A(\theta_0)-2(1-q_n)A(\theta_n^*)+2A(\mu_{n,1})\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
&E_{\mu_{n,1}}\left[b(X)-A'(\theta_n^*)\right]^2 \\
&= E_{\mu_{n,1}}[b(X)-A'(\mu_{n,1})+A'(\mu_{n,1})-A'(\theta_n^*)]^2 \\
&= A''(\mu_{n,1}) + (A'(\mu_{n,1})-A'(\theta_n^*))^2
\end{aligned}
$$

By the assumptions we obtain the right hand side of in (6.15) is upper bounded by a constant converging to zero as $n \to \infty$.

Step 3. Let $g(\underline{X},\theta) := \ddot{\psi}_n(\underline{X},\theta)$. For some fixed $\widetilde{\theta} < \overline{\overline{\theta}} < \theta_n^*$, we have that

$$
\left|g(\underline{X};\widetilde{\theta}) - g(\underline{X};\theta_n^*)\right| = \left|g'(\underline{X},\overline{\overline{\theta}})\right|\left|\widetilde{\theta}-\theta_n^*\right| \leq \sup_{\theta\in\Theta}\left|g'(\underline{X},\theta)\right|\left|\widetilde{\theta}-\theta_n^*\right|. \quad (6.16)
$$

A calculation shows that

$$
g'(\underline{X},\theta) = n^{-1}\sum_{i=1}^{n} e^{(1-q_n)\varphi}[(1-q_n)^3(\varphi')^3\varphi' + 5(1-q_n)^2(\varphi'')(\varphi')^2 \quad (6.17)
$$

$$
4(1-q_n)(\varphi''')(\varphi') + 3(1-q_n)(\varphi''')^2 + \varphi''''], \quad (6.18)
$$

where $\varphi = \varphi(X_i, \theta)$ and all the derivatives are taken with respect the parameter. Since Θ is compact, by assumptions A.3 and A.5 we have that $\sup_{\theta\in\Theta} \left|\frac{\partial^k}{\partial\theta^k} A(\theta)\right| < C_k < \infty$ for some constants C_k and $k = 1, 2, 3, 4$. Moreover,

$$e^{(1-q)\varphi(X_i,\theta)} = e^{(1-q)b(X_i)\theta} e^{-A(\theta)} \le 2(e^{(1-q)b(X_i)\theta_0} + e^{(1-q)b(X_i)\theta_1})c^*, \qquad (6.19)$$

where θ_0 and θ_1 are the boudary points in Θ and c^* is some constant such that $\sup_{\theta\in\Theta} e^{-(1-q)A(\theta)} < c^*$. Thus, given $\varepsilon > 0$, by Markov's inequality we have

$$P\left(\sup_{\theta\in\Theta} |g'(\underline{X}, \theta)| > \varepsilon\right) \le \varepsilon^{-1} E \sup_{\theta\in\Theta} |g'(\underline{X}, \theta)| < K^* < \infty, \qquad (6.20)$$

for some constant K^*. In addition, recall that in step 2 we obtained that $\dot{\psi}(\theta_n^*)$ is lower bounded by a constant. Hence, $\ddot{\psi}_n\left(\underline{X}; \widetilde{\theta}_0\right) / \dot{\psi}\left(\theta_n^*\right)$ is bounded in probability.

Since the third term in the expansion (6.13) is of higher order, by combining steps 1, 2 and 3 and applying Slutsky's Lemma we obtain the desired result.

Proof of Theorem 4.1

From the second order Taylor expansion of $\alpha(x_{L,n}; \widetilde{\theta}_n)$ about θ_n^* one can obtain

$$\sqrt{n}\frac{\left(\alpha(x_{L,n}; \widetilde{\theta}_n) - \alpha(x_{L,n}; \theta_n^*)\right)}{\sigma_n \alpha'(x_{L,n}; \theta_n^*)} \qquad (6.21)$$

$$= \sqrt{n}\frac{(\widetilde{\theta}_n - \theta_n^*)}{\sigma_n} + \frac{1}{2\sigma_n}\frac{\alpha''(x_{L,n}; \widetilde{\theta})}{\alpha'(x_{L,n}; \theta_n^*)}\sqrt{n}(\widetilde{\theta}_n - \theta_n^*)^2 \qquad (6.22)$$

$$= \sqrt{n}\frac{(\widetilde{\theta}_n - \theta_n^*)}{\sigma_n} + \frac{1}{2\sigma_n}\frac{\alpha''(x_{L,n}; \theta_n^*)}{\alpha'(x_{L,n}; \theta_n^*)}\frac{\alpha''(x_{L,n}; \widetilde{\theta})}{\alpha''(x_{L,n}; \theta_n^*)}\sqrt{n}(\widetilde{\theta}_n - \theta_n^*)^2, \qquad (6.23)$$

where $\widetilde{\theta}$ is a value between $\widetilde{\theta}_n$ and θ_n^*. Consider,

$$\alpha'(x_L; \theta) = \frac{\partial}{\partial\theta}\int_{x_L}^{\infty} e^{\varphi(x,\theta)} d\mu(x) = -\int_{-\infty}^{x_L} e^{\varphi(x,\theta)}\varphi'(x,\theta) d\mu(x),$$

and

$$\alpha''(x_L;\theta) = -\int_{-\infty}^{x_L} e^{\varphi(x,\theta)} \left(\varphi'(x,\theta)^2 + \varphi''(x,\theta)\right) d\mu(x)$$

One can see that $\alpha'(x_{L,n};\theta_n^*)/\alpha''(x_{L,n};_n^*)$ is an indetermined form of the type 0/0 as $n \to \infty$. Note that for any function $g(x, l(x))$, where $g(\cdot,\cdot)$ and $l(\cdot)$ are continuous, we have that

$$\frac{\partial}{\partial x} g(x, l(x)) = D_1 f(x, l(x)) + D_2 f(x, l(x)) \frac{\partial}{\partial x} l(x). \tag{6.24}$$

where D_1 and D_2 denote differentiation with respect the first and the second argument of $f(\cdot,\cdot)$, respectively. We apply (6.24) to $\alpha''(x_{L,n};\theta_n^*)$, obtaining

$$\begin{aligned}
&\frac{\partial}{\partial x_{L,n}} \alpha''(x_{L,n};\theta_n^*) = -\exp\left\{\varphi(x_{L,n},\theta_n^*)\right\} \left(\varphi'(x_{L,n},\theta_n^*)^2 + \varphi''(x_{L,n},\theta_n^*)\right) \\
&- \int_{-\infty}^{x_{L,n}} \exp\left\{\varphi(x,\theta_n^*)\right\} \left(\varphi'(x,\theta_n^*)^3 + 3\varphi''(x,\theta_n^*)\varphi'(x,\theta_n^*) + \varphi'''(x,\theta_n^*)\right) d\mu(x) \frac{\partial\theta_n^*}{\partial x_{L,n}}.
\end{aligned}$$

In particular, when n is large we have that

$$\begin{aligned}
&\int_{-\infty}^{x_{L,n}} \exp\left\{\varphi(x,\theta_n^*)\right\} \left(\varphi'(x,\theta_n^*)^3 + 3\varphi''(x,\theta_n^*)\varphi'(x,\theta_n^*)^2 + \varphi'''(x,\theta_n^*)\right) d\mu(x) \\
&\to A'''(\theta_n^*) - 3A''(\theta_n^*) E_{\theta_n^*}\left[b(x) - A'(\theta_n^*)\right] - A'''(\theta_n^*) = 0.
\end{aligned}$$

Further, applying (6.24) to $\alpha'(x_{L,n};\theta_n^*)$ we obtain

$$\begin{aligned}
&\frac{\partial}{\partial x_{L,n}} \alpha'(x_{L,n};\theta_n^*) = -\exp\left\{\varphi(x_{L,n},\theta_n^*)\right\} \varphi'(x_{L,n},\theta_n^*) \\
&- \int_{-\infty}^{x_{L,n}} \exp\left\{\varphi(x,\theta_n^*)\right\} \left(\varphi''(x,\theta_n^*)^2 + \varphi''(x,\theta_n^*)\right) d\mu(x) \frac{\partial\theta_n^*}{\partial x_{L,n}},
\end{aligned}$$

where,

$$\begin{aligned}
&\int_{-\infty}^{x_{L,n}} \exp\left\{\varphi(x,\theta_n^*)\right\} \left(\varphi'(x,\theta_n^*)^2 + \varphi''(x,\theta_n^*)\right) d\mu(x) \\
&\to E_{\theta_n^*}\left[\varphi(X,\theta_n^*)^2 + \varphi''(X,\theta_n^*)\right] = 0 \quad \text{as } n \to \infty.
\end{aligned}$$

A straightforward calculation shows that

$$\frac{\frac{\partial}{\partial x_{L,n}}\alpha''(x_{L,n};\theta_n^*)}{\frac{\partial}{\partial x_{L,n}}\alpha'(x_{L,n};\theta_n^*)} = \frac{(b(x_{L,n}) - A'(\theta_n^*))^2 - A''(\theta_n^*)}{b(x_{L,n}) - A'(\theta_n^*)},$$

where the right hand side of the above expression is of order $b(x_{L,n})$. Therefore, applying l'Hôpital rule we have that $\alpha''(x_{L,n};\theta_n^*)/\alpha'(x_{L,n};\theta_n^*) = o\left(n^{1/2}\right)$. The assumptions and Theorem 3.3 imply that the last term on the right hand side of eq.(6.23) is $o_P(1)$. Hence, the desired result follows from Slutsky's lemma.

Proof of Theorem 4.3

The idea of this proof is analogous to that of Theorem 4.1. From the second order Taylor expansion of $\rho(\widetilde{\theta}_n; s)$ about θ_n^* one can obtain

$$\sqrt{n}\frac{\rho(s_n;\widetilde{\theta}_n) - \rho(s_n;\theta_n^*)}{\sigma_n \rho'(s_n;\theta_n^*)} = \sqrt{n}\frac{(\widetilde{\theta}_n - \theta_n^*)}{\sigma_n} + \frac{1}{2\sigma_n}\frac{\rho''(s_n;\bar{\bar{\theta}})}{\rho'(s_n;\theta_n^*)}\sqrt{n}(\widetilde{\theta}_n - \theta_n^*)^2. \quad (6.25)$$

where $\bar{\bar{\theta}}$ is a value between $\widetilde{\theta}_n$ and θ_n^* and ρ', ρ'' denote first and second derivatives with respect the parameter. It is clear that assumption (ii) implies that $\rho''(s_n;\bar{\bar{\theta}})/\rho'(s_n;\theta_n^*) = o(n^{1/2})$ and thus the desired result follows from Slutsky's Lemma combined with theorem 4.1.

Suppose instead that $\rho''(s_n;\bar{\bar{\theta}})/\rho'(s_n;\theta_n^*)$ is an indeterminate form. Write

$$\frac{\rho''(s_n;\bar{\bar{\theta}})}{\rho'(s_n;\theta_n^*)} = \frac{\rho''(s_n;\theta_n^*)}{\rho'(s_n;\theta_n^*)}\frac{\rho''(s_n;\bar{\bar{\theta}})}{\rho''(s_n;\theta_n^*)} \quad (6.26)$$

Note that $s = \rho^{-1}(\rho(s))$. Differentiating with respect s and using derivatives' the chain rule, we obtain $1 = \frac{\partial \rho^{-1}(t)}{\partial t}\frac{\partial \rho(s)}{\partial s}$. In particular, we have that $\frac{\partial \rho(s;\theta)}{\partial s} = e^{-b(\rho(s))\theta + A(\theta)}$. Thus, we can compute

$$K(s;\theta) = \frac{\frac{\partial}{\partial s}\frac{\partial^2}{\partial \theta^2}\rho(s;\theta)}{\frac{\partial}{\partial s}\frac{\partial}{\partial \theta}\rho(s;\theta)} = \frac{e^{\theta b(\rho(s)) - A(\theta))}[(A'(\theta) - b(\rho(s)))^2 + A''(\theta)]}{e^{\theta b(\rho(s)) - A(\theta))}(A'(\theta) - b(\rho(s)))},$$

where the last equality is obtained by switching the order of the derivatives, since $\rho(s;\theta)$ and $\frac{\partial}{\partial\theta}\rho(s;\theta)$ are continuous in both of the arguments. Moreover, note that $K(s_n;\theta)$ is of order $b\left(\rho(s_n;\theta)\right)$. Therefore, by l'Hôpital rule we have that $\rho''(s_n;\theta_n^*)/\rho'(s_n;\theta_n^*)$ is of order $b\left(\rho(s_n;\theta)\right)$. Finally, Theorem 4.1 combined with assumption (i) imply that the last term in eq.(6.25) is $o_P(1)$. The desired result is obtained by applying Slutsky's Lemma.

Appendix B - Calculation of the asymptotic variance of the MLqE for an exponential distribution

First, consider the equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial\lambda}\int_0^\infty L_q\left(f(x;\lambda)\right) f(x;\lambda_0)dx \\ &= \lambda^{-q}\int_0^\infty (1-\lambda x)e^{-[\lambda(1-q)+\lambda_0]x}dx \\ &= \lambda^{-q}\lambda_0\left[\frac{1}{\lambda(1-q)+\lambda_0}-\frac{\lambda}{\left(\lambda(1-q)+\lambda_0\right)^2}\right] \end{aligned}$$

and the solution of the the above equation can be easily computed as $\lambda^* = \lambda_0/q$. Next consider,

$$E_{\lambda_0}\left[\frac{\partial}{\partial\lambda}L_q f(x;\lambda)\right]^2 = \lambda_0\lambda^{-2q}\int_0^\infty (1-2\lambda x+\lambda^2x^2)e^{-[2\lambda(1-q)+\lambda_0]x}dx$$

Computing the integrals and setting $\lambda=\lambda^*$ gives

$$\begin{aligned} E_{\lambda_0}\left[\frac{f'(x;\lambda^*)}{f(x;\lambda^*)^q}\right]^2 &= \lambda_0\left(\frac{\lambda_0}{q}\right)^{-2q}\left[\frac{1}{\lambda_0(\frac{2}{q}-1)}-\frac{2(\lambda_0/q)}{\lambda_0^2(\frac{2}{q}-1)^2}+\frac{2(\lambda_0/q)^2}{\lambda_0^3(\frac{2}{q}-1)^3}\right] \\ &= q\left(\frac{\lambda_0}{q}\right)^{-2q}\left[\frac{q^2-2q+2}{(2-q)^3}\right]. \end{aligned}$$

Next compute

$$E_{\lambda_0}\left[\frac{\partial}{\partial\lambda}\frac{f'(x;\lambda)}{f(x;\lambda)^q}\right] = \lambda_0\lambda^{-q}\frac{\partial}{\partial\lambda}\left[\int_0^\infty (1-\lambda x)e^{-[\lambda(1-q)+\lambda_0]x}dx\right]$$
$$= \frac{\partial}{\partial\lambda}\left[\frac{\lambda_0\lambda^{-q}}{\lambda(1-q)+\lambda_0} - \frac{\lambda^{1-q}\lambda_0}{[\lambda(1-q)+\lambda_0]^2}\right]$$
$$= 2\lambda_0(1-q)\frac{\lambda^{1-q}}{(\lambda_0+\lambda(1-q))^3} - 2\frac{\lambda_0}{\lambda^q}\frac{1-q}{(\lambda_0+\lambda(1-q))^2}$$
$$- q\frac{\lambda_0}{\lambda^{q+1}(\lambda_0+\lambda(1-q))}.$$

Substituting for $\lambda = \lambda^*$ in the above expression and reorganizing gives:

$$E_{\lambda_0}\left[\frac{\partial}{\partial\lambda}\frac{f'(x;\lambda)}{f(x;\lambda)^q}\bigg|_{\lambda=\lambda^*}\right] = -\frac{q^2}{\lambda_0(\lambda_0/q)^q}.$$

Finally, the asymptotic variance is obtained as

$$\sigma^2(\lambda^*) = \frac{E_{\lambda_0}\left[\frac{f'(x;\lambda^*)}{f(x;\lambda^*)^q}\right]^2}{\left(E_{\lambda_0}\frac{\partial}{\partial\lambda}\frac{f'(x;\lambda)}{f(x;\lambda)^q}\Big|_{\lambda=\lambda^*}\right)^2} = \left(\frac{\lambda_0}{q_n}\right)^2\left[\frac{q^2-2q+2}{q^3(2-q)^3}\right].$$

References

[1] J. Aczél and Z. Daroczy. *On Measures of Information and Their Characterization.* Academic Press, New York, 1975.

[2] A.W. Van der Vaart. *Asymptotic Statistics.* Cambridge University Press, New York, 1998.

[3] T. S. Ferguson. *A Course in Large Sample Theory.* Chapman & Hall Ltd, 1996.

[4] Murray Gell-Mann, editor. *Nonextensive Entropy, Interdisciplinary Applications.* Number 3. Oxford University Press, New York, 2004.

[5] D Goldfarb. A family of variable-metric method derived by variational means. *Mathematics of Computation*, 24:23–26, 1970.

[6] J Havrda and F Charvat. *Kibernetika*, 3(30), 1967.

[7] Feifang Hu and James V. Zidek. *The Relevance Weighted Likelihood with Applications.* Springer-Verlag Inc., Berlin; New York, 2001.

[8] Heifang Hu and James V. Zidek. The weighted likelihood. *The Canadian Journal of Statistics*, 30(3):347–371, Sept 2002.

[9] P J Huber. *Robust Statistics.* Wiley Series in Probability. John Wiley and Sons, 1981.

[10] S Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[11] S Kullback and R A Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[12] Alfréd Rényi. *Selected Papers of Alfréd Rényi.* Akademia Kiado, Budapest, 1969.

[13] C E Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[14] Robert Tibshirani and Trevor Hastie. Local likelihood estimation. *Journal of American Statistical Association*, 82(398):559–576, June 1987.

[15] Constantino Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52(1-2):479–487, July 1988.

[16] Xiaogang Wang and James V. Zidek. Selecting likelihood weights by cross-validation. *The Annals of Statistics*, 33(2):463–500, 2005.

[17] Constance van Eeden Xiaogang Wang and James V. Zidek. Asympotic properties of maximum weighted likelihood estimators. *Journal of Inference and Statistical planning*, 119:37–54, 2004.